An International Journal
computers & mathematics with applications

PERGAMON Computers and Mathematics with Applications 43 (2002) 657–669
www.elsevier.com/locate/camwa

A New Measure of Irregularity of Distribution and Quasi-Monte Carlo Methods for Global Optimization

XIAOQUN WANG
Department of Mathematical Sciences
Tsinghua University
Beijing 100084, P.R. China
xwang@math.tsinghua.edu.cn

(Received May 2000; revised and accepted May 2001)

Abstract—Measures of irregularity of distribution, such as discrepancy and dispersion, play a major role in quasi-Monte Carlo methods for integration and optimization. In this paper, a new measure of irregularity of distribution, called *volume-dispersion*, is introduced. Its relation to the discrepancy and traditional dispersion, and its applications in global optimization problems are investigated. Optimization errors are bounded in terms of the volume-dispersion. Also, the volume-dispersion is generalized to the so-called *F-volume-dispersion* and *quasi-F-volume-dispersion*. They are reasonable measures of representation of point sets for given probability distributions on general domains and have potential applications in optimization problems when prior knowledge about the possible location of the optimizer is known and in the problems of experimental designs. Methods of generating point sets with low quasi-F-volume-dispersion are described.

Keywords—Quasi-Monte Carlo methods, Irregularity of distribution, Dispersion, Discrepancy, Optimization.

1. INTRODUCTION

Quasi-Monte Carlo (QMC) methods can be described as deterministic versions of Monte Carlo methods, in the sense that the random samples in Monte Carlo methods are replaced by well-chosen deterministic points. Particular interest in QMC methods has centered on measures of irregularity or uniformity of distribution, and on the development of sequences well distributed according to such measures. Discrepancy and dispersion, the two well-known measures of irregularity of distribution of point sets, play a major role in QMC methods for choosing good point sets. The discrepancy criterion is particularly suited for numerical integration, and the dispersion criterion is especially useful in studying global optimization problems [1,2].

We recall the definitions of discrepancies and dispersion. Let $\mathcal{P}_N = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a point set of N points in the s-dimensional unit cube $I^s = [0,1]^s$ and let $A(J; \mathcal{P}_N)$ denote the number of points of $\mathcal{P}_N$ falling in the region J.

Supported by the NSF of China under Grants 79970120 and 10001021.
The author would like to thank the referees for their valuable comments and suggestions.

0898-1221/02/$ - see front matter Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX
PII: S0898-1221(01)00311-X

DEFINITION 1.1. *The (extreme) discrepancy $D(\mathcal{P}_N)$ of $\mathcal{P}_N$ is defined by*

$$D(\mathcal{P}_N) = \sup_{J \in E} \left| \frac{A(J; \mathcal{P}_N)}{N} - \mathrm{Vol}\,(J) \right|, \tag{1}$$

where E is the family of all subintervals of I^s of the form $\prod_{i=1}^{s}[u_i, v_i]$ and $\mathrm{Vol}\,(J)$ is the volume of the subinterval J. If E is the family of all subintervals of I^s of the form $\prod_{i=1}^{s}[0, v_i]$, then the star discrepancy is obtained.

The importance of discrepancy for QMC integration can be seen from the well-known Koksma-Hlawka inequality (see [1]). There are several methods for constructing low discrepancy sequences for which the discrepancy is bounded by a constant times $N^{-1}(\log N)^s$, which suggests much greater uniformity than a sequence of random numbers [1,3–5].

DEFINITION 1.2. *Letting ρ be a metric on I^s, the dispersion $d(\mathcal{P}_N, \rho)$ of $\mathcal{P}_N$ is defined by*

$$d(\mathcal{P}_N, \rho) = \sup_{\mathbf{x} \in I^s} \min_{1 \le n \le N} \rho(\mathbf{x}, \mathbf{x}_n). \tag{2}$$

The dispersion criterion is regarded as more significant than discrepancy in global optimization theory. Consider the problem of global optimization

$$f^* = \sup_{\mathbf{x} \in I^s} f(\mathbf{x}), \tag{3}$$

where $f : I^s \to \mathbf{R}$ is a function defined on the s-dimensional unit cube I^s.

There are many gradient methods for this kind of problem. But if the objective function is not unimodal and the dimension s is large, it is often difficult to reach the global maximum (often the local maximum is obtained), since the solution often depends on the choice of the initial point. Therefore, we sometimes use *QMC search algorithms* (also known as *quasi-random search methods*), which are deterministic analogs of random search procedures [6] for approximating global extrema of a function. To approximate f^*, suppose that we are allowed to choose N points $\mathbf{x}_1, \dots, \mathbf{x}_N$ in I^s at which to observe the values of the function f. Letting $f_N^* = \max_{1 \le i \le N} f(\mathbf{x}_i)$, then the approximate error can be bounded in terms of the dispersion $d(\mathcal{P}_N, \rho)$ (see [1,7])

$$f^* - f_N^* \le w(f, d(\mathcal{P}_N, \rho)), \tag{4}$$

where $w(f, t)$ is the *modulus of continuity* of $f(\mathbf{x})$, defined by

$$w(f, t) = \sup_{\substack{\rho(\mathbf{u}, \mathbf{v}) \le t \\ \mathbf{u}, \mathbf{v} \in I^s}} |f(\mathbf{u}) - f(\mathbf{v})|, \qquad \text{for } t \ge 0.$$

If f is continuous on I^s, then the convergence to the global maximum is assured, whenever $d(\mathcal{P}_N, \rho) \to 0$ as $N \to \infty$. The goal of QMC search is to choose the points in such a way that f_N^* is a good approximation to f^*. The inequality (4) indicates the usefulness of point sets with small dispersion in QMC search algorithms.

This paper is organized as follows. In Section 2, motivated by an equivalent definition of the traditional dispersion, a new measure of irregularity of distribution, called *volume-dispersion*, is proposed. Its relation to the discrepancy and traditional dispersion, and its applications in optimization problems are investigated. In Section 3, we extend the concept of volume-dispersion and define the so-called *F-volume-dispersion* and *quasi-F-volume-dispersion*. They are reasonable measures of representation of point sets for given probability distribution. Such a distribution might represent prior knowledge about the likely location of the optimizer. Finally, we conclude and present directions for future research in Section 4.

2. A NEW MEASURE OF IRREGULARITY OF DISTRIBUTION

2.1. The Definition of the New Measure—Volume-Dispersion

We begin by giving an equivalent definition of the traditional dispersion defined in (2). Let the point set $\mathcal{P}_N$ and the metric ρ be the same as in Definition 1.2. Let

$$B(\mathbf{x}, r, \rho) = \{\mathbf{z} \in I^s : \rho(\mathbf{z}, \mathbf{x}) \le r\}.$$

THEOREM 2.1. *The following is an equivalent definition of the dispersion $d(\mathcal{P}_N, \rho)$:*

$$d(\mathcal{P}_N, \rho) = \sup_{\substack{\mathbf{x}\in I^s \\ r\ge 0}} \{r : B(\mathbf{x}, r, \rho) \cap \mathcal{P}_N = \emptyset\}, \tag{5}$$

where the supremum is taken over all the points $\mathbf{x} \in I^s$ and all radii $r \ge 0$. In other words, the dispersion of $\mathcal{P}_N$ is the maximum radius of the ball which contains no point of $\mathcal{P}_N$.

PROOF. Denote the quantities defined in (2) and (5) by d_1 and d_2, respectively. We must prove that $d_1 = d_2$.

By the definition of d_1, for arbitrary ε with $0 < \varepsilon < d_1$, there exists a point $\mathbf{x}^* \in I^s$, such that $\rho(\mathbf{x}^*, \mathbf{x}_n) > d_1 - \varepsilon$ for $1 \le n \le N$. This means that the ball $B(\mathbf{x}^*, d_1 - \varepsilon, \rho)$ contains none of the points $\mathbf{x}_1, \dots, \mathbf{x}_N$, i.e., $B(\mathbf{x}^*, d_1 - \varepsilon, \rho) \cap \mathcal{P}_N = \emptyset$. Hence, $d_2 \ge d_1 - \varepsilon$. This implies $d_2 \ge d_1$.

On the other hand, by the definition of d_2, for arbitrary ε with $0 < \varepsilon < d_2$, there exists a point $\mathbf{x}_* \in I^s$, such that

$$\sup_{r\ge 0}\{r : B(\mathbf{x}_*, r, \rho) \cap \mathcal{P}_N = \emptyset\} > d_2 - \varepsilon.$$

Hence, $B(\mathbf{x}_*, d_2 - \varepsilon, \rho) \cap \mathcal{P}_N = \emptyset$. This means $\rho(\mathbf{x}_*, \mathbf{x}_n) \ge d_2 - \varepsilon$ for $1 \le n \le N$. Thus, $d_1 \ge \min_{1\le n\le N} \rho(\mathbf{x}_*, \mathbf{x}_n) \ge d_2 - \varepsilon$. This implies $d_1 \ge d_2$. This completes the proof. ∎

Motivated by Theorem 2.1, if we consider the "volume" (i.e., Lebesgue measure) of the ball $B(\mathbf{x}, r, \rho)$ instead of its radius r, then the so-called "volume-dispersion" can be defined.

DEFINITION 2.1. *The volume-dispersion $\mathcal{V}_d(\mathcal{P}_N, \rho)$ of the point set $\mathcal{P}_N$ is defined by*

$$\mathcal{V}_d(\mathcal{P}_N, \rho) = \sup_{\substack{\mathbf{x}\in I^s \\ r\ge 0}} \{\mathrm{Vol}\,[B(\mathbf{x}, r, \rho)] : B(\mathbf{x}, r, \rho) \cap \mathcal{P}_N = \emptyset\},$$

where the supremum is taken over all $\mathbf{x} \in I^s$ and all radii $r \ge 0$, and $\mathrm{Vol}\,[B(\mathbf{x}, r, \rho)]$ denotes the volume (Lebesgue measure) of the ball $B(\mathbf{x}, r, \rho)$. If the metric ρ can be understood from the context, we simply write $\mathcal{V}_d(\mathcal{P}_N)$ for the volume-dispersion.

Intuitively, the volume-dispersion of $\mathcal{P}_N$ is the largest volume of the ball $B(\mathbf{x}, r, \rho)$, which contains no point of $\mathcal{P}_N$. If the volume-dispersion is small, then the set $\mathcal{P}_N$ leaves no big "empty" ball. In this sense, the volume-dispersion is a reasonable measure of irregularity of distribution of a point set.

The volume-dispersion depends on the metric. Two metrics are common: the standard Euclidean metric ρ_2 and the maximum metric ρ_∞. The maximum metric is defined as: $\rho_\infty(\mathbf{x}, \mathbf{y}) = \max_{1\le i\le s} |x_i - y_i|$, for $\mathbf{x} = (x_1, \dots, x_s)$ and $\mathbf{y} = (y_1, \dots, y_s)$. The relation between the discrepancy and dispersion is well established under these metrics. Under the standard Euclidean metric ρ_2, one has $d(\mathcal{P}_N, \rho_2) \le \sqrt{s} D(\mathcal{P}_N)^{1/s}$. Under the maximum metric ρ_∞, one has $d(\mathcal{P}_N, \rho_\infty) \le D(\mathcal{P}_N)^{1/s}$ (see [1]).

Since the two metrics are related $\rho_\infty(\mathbf{x}, \mathbf{y}) \le \rho_2(\mathbf{x}, \mathbf{y}) \le s^{1/2} \rho_\infty(\mathbf{x}, \mathbf{y})$, the volume-dispersions associating these two metrics have the same order of magnitude. In the following, we investigate the relation of volume-dispersion with the discrepancy and traditional dispersion under the maximum metric. We will also study the weighted metric, where the weights represent the different dependence of the function f on various input variables.

2.2. Volume-Dispersion under the Maximum Metric

Volume-dispersion has a close connection to the discrepancy as shown in the following. Such a connection has a very simple form under the maximum metric.

THEOREM 2.2. *Letting $\mathcal{P}_N$ be a point set of N points in I^s, then the volume-dispersion of $\mathcal{P}_N$ with respect to the maximum metric ρ_∞ is bounded by its discrepancy $D(\mathcal{P}_N)$, i.e.,*

$$\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \le D(\mathcal{P}_N).$$

Theorem 2.2 corresponds to a special case of a more general result in Section 3 (see Theorem 3.1, where a more general result is proved). From this theorem, the notion of volume-dispersion is weaker than that of discrepancy. Any low discrepancy point set (or sequence) is a low volume-dispersion point set (or sequence), but not conversely. For example, the sequence defined by $x_1 = 1$, $x_n = \{\log(2n-3)/\log 2\}$, for $n \ge 2$, satisfies

$$\lim_{N\to\infty} N\mathcal{V}_d(\mathcal{P}_N) = \frac{1}{\log 2},$$

where $\mathcal{P}_N$ denotes the first N terms of the sequence, but $\lim_{N\to\infty} D(\mathcal{P}_N) \neq 0$.

The next theorem establishes the relation of the volume-dispersion with the traditional dispersion.

THEOREM 2.3. *Letting $\mathcal{P}_N$ be a point set of N points in I^s, then*

$$[d(\mathcal{P}_N, \rho_\infty)]^s \le \mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \le 2^s [d(\mathcal{P}_N, \rho_\infty)]^s. \tag{6}$$

PROOF. Let $r = d(\mathcal{P}_N, \rho_\infty) > 0$. For arbitrary ε with $0 < \varepsilon < r$, there exists a point $\mathbf{x}^* \in I^s$, such that the ball $B(\mathbf{x}^*, r-\varepsilon, \rho_\infty)$ contains none of the points $\mathbf{x}_1, \dots, \mathbf{x}_N$, i.e., $B(\mathbf{x}^*, r-\varepsilon, \rho_\infty) \cap \mathcal{P}_N = \emptyset$. Note that $J = B(\mathbf{x}^*, r-\varepsilon, \rho_\infty)$ is a closed ball with volume $\mathrm{Vol}\,(J) \ge (r-\varepsilon)^s$. So

$$\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \ge \mathrm{Vol}\,(J) \ge (r-\varepsilon)^s.$$

This implies that $\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \ge [d(\mathcal{P}_N, \rho_\infty)]^s$.

The second part of relation (6) follows from the fact that for any ball $B(\mathbf{x}, r, \rho_\infty)$, its volume is bounded by 2^s times its radius r. ∎

Theorem 2.3 indicates that any low dispersion point set is a low volume-dispersion point set, and vice versa. Therefore, we could expect that low volume-dispersion point sets play a similar role in QMC search algorithms as low dispersion point sets do. In fact, the optimization error can be bounded in terms of the volume-dispersion.

THEOREM 2.4. *Under the same notations as in (4), we have*

$$f^* - f_N^* \le w\left(f, [\mathcal{V}_d(\mathcal{P}_N, \rho_\infty)]^{1/s}\right),$$

where $w(f,t)$ is the modulus of continuity of $f(x)$ under the metric ρ_∞.

If the function f satisfies a Lipschitz condition, i.e., $|f(\mathbf{x}) - f(\mathbf{y})| \le L\rho_\infty(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y}$ in I^s, where L is a constant, then

$$f^* - f_N^* \le L\,[\mathcal{V}_d(\mathcal{P}_N, \rho_\infty)]^{1/s}.$$

The first inequality in this theorem follows from inequality (4) and Theorem 2.3. The second one is a special case of the first one, since the modulus of continuity of $f(\mathbf{x})$ can be bounded as: $w(f,t) \le L\,t$, for any f satisfying the Lipschitz condition.

The lower bound on traditional dispersion is known: $d(\mathcal{P}_N, \rho_\infty) \ge 2^{-1}N^{-1/s}$ (see [1]). Based on this and Theorem 2.3, we have a lower bound for the volume-dispersion.

COROLLARY 2.5. *For any point set $\mathcal{P}_N$ consisting of N points in I^s, we have*

$$\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \geq 2^{-s} N^{-1}.$$

Therefore, the volume-dispersion of a point set is at least of the order of the magnitude N^{-1}. On the other hand, we will see that for the Hammersley point set, the Halton sequence, the (t, m, s)-net, and the (t, s)-sequence, their volume-dispersions attain this order of magnitude. For details about these point sets and sequences, we refer to [1]. Since the bounds for the traditional dispersion of these point sets and sequences are established (see [1]), according to Theorem 2.3, we have the following upper bounds for their volume-dispersions.

THEOREM 2.6.

(1) *For the N-element Hammersley point set $\mathcal{P}_N$ in the pairwise relatively prime bases $b_1, \dots, b_{s-1}$, we have*

$$\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \leq 2^s \left(1 + \max_{1 \leq i \leq s-1} b_i\right)^s N^{-1}.$$

(2) *For the Halton sequence in the pairwise relatively prime bases $b_1, \dots, b_s$,*

$$\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \leq 2^s \left(\max_{1 \leq i \leq s} b_i^s\right) N^{-1},$$

where $\mathcal{P}_N$ is the set consisting of the first N terms of the Halton sequence.

(3) *For any (t, m, s)-net $\mathcal{P}_N$ in base b, we have*

$$\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \leq 2^s b^{s+t-1} N^{-1}, \qquad \text{with } N = b^m.$$

(4) *For any (t, s)-sequence in base b, we have*

$$\mathcal{V}_d(\mathcal{P}_N, \rho_\infty) \leq 2^s b^{s+t} N^{-1},$$

where $\mathcal{P}_N$ is the set consisting of the first N terms of the (t, s)- sequence.

This theorem shows that the Hammersley point set, the Halton sequence, the (t, m, s)-net, and the (t, s)-sequence have the lowest possible order of volume-dispersion. The convergence rate of QMC search algorithms using these point sets or sequences will be $O(N^{-1/s})$ according to Theorem 2.4.

2.3. The Computation of the Volume-Dispersion

How shall we compute the volume-dispersion of a point set? In one-dimension, a simple explicit formula can be derived.

THEOREM 2.7.

(1) *If the one-dimensional points $x_1, \dots, x_N$ of $\mathcal{P}_N$ are ordered such that $0 \leq x_1 \leq x_2 \leq \cdots \leq x_N \leq 1$, then*

$$\mathcal{V}_d(\mathcal{P}_N) = \max(x_1, x_2 - x_1, \dots, x_N - x_{N-1}, 1 - x_N).$$

(2) *For any point set $\mathcal{P}_N$ consisting of N points in $[0, 1]$,*

$$\mathcal{V}_d(\mathcal{P}_N) \geq \frac{1}{N+1},$$

and this bound is achieved by the point set $P_N^{(1)} = \{i/(N+1),\, i = 1, \dots, N\}$.

For comparison, we note that the traditional dispersion of the point set $\mathcal{P}_N = \{x_1, \dots, x_N\}$ with $0 \le x_1 \le \cdots \le x_N \le 1$ is given by (see [1])

$$d(\mathcal{P}_N) = \max\left(x_1, \frac{1}{2}(x_2 - x_1), \dots, \frac{1}{2}(x_N - x_{N-1}), 1 - x_N\right).$$

In dimension $s \ge 2$, the exact value of volume-dispersion is not easy to obtain in general. We have not yet got a formula for an arbitrary point set. This is a topic of ongoing research.

When the distribution of the points of $\mathcal{P}_N$ is regular in I^s, we can compute its volume-dispersion directly. As an example, we compute the volume-dispersion of the following *equi-lattice points* in I^s. For comparison, we also compute its traditional dispersion.

EXAMPLE. Let m be a positive integer. Consider the point set of $N = m^s$ points

$$P^* = \left\{\left(\frac{2k_1 - 1}{2m}, \dots, \frac{2k_s - 1}{2m}\right),\ k_i = 1, \dots, m;\ i = 1, \dots, s\right\}.$$

The volume-dispersion of P^* is equal to the volume of the largest empty ball. Such a ball can be easily seen to be an s-dimensional cube with edge-length $1/m$, and thus,

$$\mathcal{V}_d(P^*, \rho_\infty) = \left(\frac{1}{m}\right)^s = N^{-1}.$$

The traditional dispersion is equal to the radius of the largest empty ball, so

$$d(P^*, \rho_\infty) = \frac{1}{2m} = \frac{1}{2} N^{-1/s}.$$

This example shows that the volume-dispersion of equi-lattice point set P^* also attains the minimal order of magnitude as the well-known low discrepancy point sets and sequences do. But theoretical and numerical studies have shown that the goodness of the equi-lattice point set for global optimization (and for numerical integration and experimental design) is questionable in large dimensions (for example, $s > 5$). A closer investigation for the equi-lattice point set P^* and low discrepancy sequences shows a big difference between them. Consider their lower-dimensional projections. For the equi-lattice point set P^*, let $P^*(i_1, \dots, i_t)$ denote its projection onto the dimensions $i_1, \dots, i_t$, where $1 \le t \le s$. Then the point set $P^*(i_1, \dots, i_t)$ only contains m^t different points. Its volume-dispersion is

$$d(P^*(i_1, \dots, i_t), \rho_\infty) = \left(\frac{1}{m}\right)^t = N^{-t/s},$$

which is much worse than the optimal order $O(N^{-1})$ for small value of t. Computing f at the $N = m^s$ points of P^*, we obtain only m^t essentially different values, which means a "loss of information". The situation is completely different for low discrepancy sequences. For Halton sequence, Sobol sequence, and Niederreiter sequence, their t-dimensional projections are still low discrepancy sequences (of t dimension), and their volume-dispersions have the order of magnitude of $O(N^{-1})$, which is independent of t and is also optimal in dimension t. Better projection properties are needed when the function only depends on a subset of the input variables.

2.4. Volume-Dispersion under a Weighted Metric

In many problems of numerical mathematics, the variables of the function $f(\mathbf{x})$ are not equally important; some variables are far more important than others. In multivariate numerical integration, Sloan and Woźniakowski [8] used this fact to obtain a possible theoretical explanation for the surprisingly good performance of QMC for the evaluation of high-dimensional integrals, such as those arising in finance [9].

Similar situations are encountered in the problems of optimization [10]. For functions equally dependent on all dimensions, the convergence rate of QMC optimization will be only $O(N^{-1/s})$, which is slow for large s. But functions f met in practice are often heavily dependent on a few dimensions, say t and $t \ll s$, and not very sensitive at all to others. We will show that in such situations, the convergence rate may be much better, even $O(N^{-1/t})$.

Let $l_i \geq 0$, $i = 1, \dots, s$, be a sequence of weights, which model the dependence of an objective function on various dimensions. Define a weighted metric

$$\rho_w(\mathbf{x}, \mathbf{y}) = \max_{i=1,\dots,s} [l_i |x_i - y_i|],$$

for points $\mathbf{x} = (x_1, \dots, x_s)$ and $\mathbf{y} = (y_1, \dots, y_s)$ in I^s. The weight l_i measures the importance of f on the i^{th} variable x_i. If f has bounded partial derivatives on I^s, we can use for l_i the supremum of $|\frac{\partial f}{\partial x_i}|$ over I^s. If $l_i = 1$ for all i, the weighted metric reduces to the maximum metric.

Note that under the weighted metric ρ_w, the ball $B(\mathbf{x}, r, \rho_w)$ is still a subinterval of I^s of the form $\prod_{i=1}^{s} [u_i, v_i]$. The following theorem is parallel to Theorem 2.2.

THEOREM 2.8. *The volume-dispersion of $\mathcal{P}_N$ with respect to the weighted metric ρ_w is bounded by its discrepancy $D(\mathcal{P}_N)$, i.e.,*

$$\mathcal{V}_d(\mathcal{P}_N, \rho_w) \leq D(\mathcal{P}_N).$$

Several different types of weights will be considered below. The first type is that all weights are positive. The second type is that some weights are equal to zero. The case that all weights are positive, but some of them are very small (but not zero) will also be discussed.

THEOREM 2.9. *If the weights l_i satisfy*

$$l_1 \geq l_2 \geq \cdots \geq l_t > 0, \qquad \text{but } l_{t+1} = \cdots l_s = 0 \tag{7}$$

($t = s$ corresponds to the case that all weights are positive), then

$$\frac{1}{l_1 \cdots l_t} [d(\mathcal{P}_N, \rho_w)]^t \leq \mathcal{V}_d(\mathcal{P}_N, \rho_w) \leq \frac{2^t}{l_1 \cdots l_t} [d(\mathcal{P}_N, \rho_w)]^t. \tag{8}$$

PROOF. The proof is similar to that of Theorem 2.3. Let $r = d(\mathcal{P}_N, \rho_w) > 0$. For arbitrary ε with $0 < \varepsilon < r$, there exists a point $\mathbf{x}^* \in I^s$, such that the ball $B(\mathbf{x}^*, r - \varepsilon, \rho_w)$ contains none of the points $\mathbf{x}_1, \dots, \mathbf{x}_N$, i.e., $B(\mathbf{x}^*, r - \varepsilon, \rho_w) \cap \mathcal{P}_N = \emptyset$. Note that in metric ρ_w the ball $J = B(\mathbf{x}^*, r - \varepsilon, \rho_w)$ has the form

$$B(\mathbf{x}^*, r - \varepsilon, \rho_w) = \{\mathbf{x} \in I^s : l_i |x_i - x_i^*| \leq r - \varepsilon,\ i = 1, \dots, t;\ x_j \in [0, 1],\ j = t + 1, \dots, s\}.$$

Its volume satisfies

$$\operatorname{Vol}(J) \geq \frac{1}{l_1 \cdots l_t} (r - \varepsilon)^t.$$

Thus,

$$\mathcal{V}_d(\mathcal{P}_N, \rho_w) \geq \operatorname{Vol}(J) \geq \frac{1}{l_1 \cdots l_t} (r - \varepsilon)^t.$$

This implies that

$$\mathcal{V}_d(\mathcal{P}_N, \rho_w) \geq \frac{1}{l_1 \cdots l_t} [d(\mathcal{P}_N, \rho_w)]^t.$$

The second part of the relation (8) follows from the fact that the volume-dispersion of the ball $B(\mathbf{x}, r, \rho_w)$ is bounded by $2^t / l_1 \cdots l_t$ times its radius r. ∎

The usefulness of the volume-dispersion with respect to the weighted metric in QMC search algorithms can be seen from the following theorem.

THEOREM 2.10. *Assume that the weights l_i satisfy condition (7). If f satisfies a generalized Lipschitz condition with respect to ρ_w, i.e., $|f(\mathbf{x}) - f(\mathbf{y})| \leq L \rho_w(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y}$ in I^s, where L is a constant, then*

$$f^* - f_N^* \leq L [l_1 \cdots l_t \mathcal{V}_d(\mathcal{P}_N, \rho_w)]^{1/t}. \tag{9}$$

If the point set $\mathcal{P}_N$ is one of the low discrepancy point sets in Theorem 2.6, then the convergence rate is $O(N^{-1/t})$.

REMARK. A more realistic assumption on the weights than assumption (7) is that all weights are positive, but some weights are small (but not zero). Correspondingly, one may assume that the function f depends on *all* s variables, but has the form $f = g + h$ where $g \gg h$ and g depends on t variables with $t < s$. For a given point set $\mathcal{P}_N$, if the weights satisfy

$$l_1 \geq l_2 \geq \cdots \geq l_t > 0, \quad \text{and} \quad 0 < l_j < d(\mathcal{P}_N, \rho_w), \qquad j = t+1, \ldots, s, \tag{10}$$

then we can prove that the same relations (8) and (9) hold for this case. The accuracy of QMC search depends actually on the volume-dispersion of the projection of $\mathcal{P}_N$ on the first t dimensions.

At the end of this section, we point out that the newly defined volume-dispersion has some advantages over discrepancy and traditional dispersion. An advantage of volume-dispersion over discrepancy is that volume-dispersion is independent of the coordinate axis framework. Moreover, it is *invariant* under reflection about the plane $x_j = 1/2$ passing through the center of the cube. Note that the star discrepancy lacks this property, since star discrepancy favors the origin of the unit cube.

The traditional dispersion is only defined as the measure of the derivation from the uniform distribution, and QMC search algorithms based on the traditional dispersion criterion do not take into account the prior information about the possible location of the optimizer. An advantage of the volume-dispersion over traditional dispersion is that volume-dispersion can be easily extended to general cases as measures of representation of point sets (or sequences) for arbitrary given probability distribution on general domain and such a distribution might represent the prior information about the possible location of the optimizer in the problems of optimization. Thus, the volume-dispersion can be useful for a wider class of problems. This is one of the reasons why we introduce the volume-dispersion. The next section deals with the extensions of volume-dispersion.

3. THE GENERALIZATIONS OF VOLUME-DISPERSION

If prior information about the likely location of the optimal point is available, then one should prefer nonuniform location of search points to uniform ones in searching algorithms. Such information can be obtained from an initial crude search and from the properties of the objective functions. For optimization problems with constraints, one possible way is to consider the constraints as a compact manifold embedded in some Euclidean space and use transformation of this manifold to the unit cube. The uniform distribution on the manifold will be transformed to nonuniform distribution on the unit cube. These facts indicate the necessity of measures of representation of point sets for any given probability distribution.

It is very natural to generalize the notion of volume-dispersion to nonuniform distribution on a general domain. In this section, the concepts of F-volume-dispersion and quasi-F-volume-dispersion are introduced. These concepts appear to be parallel to the concepts of F-discrepancy and quasi-F-discrepancy [2]. Note that the traditional dispersion has no counterpart for general probability distributions.

Let ρ be a metric on $\mathbf{R}^s$ and $\mathcal{P}_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a point set of N points in $\mathbf{R}^s$. Let $\mathbf{X}$ be a random vector defined on $\mathbf{R}^s$ with the cumulative distribution function (c.d.f.) $F(\mathbf{x})$. If we consider the "probability measure" instead of the "volume" of the ball which contains no point of $\mathcal{P}_N$, then the "F-volume-dispersion" can be defined.

DEFINITION 3.1. *The F-volume-dispersion $\mathcal{V}_F(\mathcal{P}_N, \rho)$ of the point set $\mathcal{P}_N$ with respect to the probability distribution $F(\mathbf{x})$ is defined by*

$$\mathcal{V}_F(\mathcal{P}_N, \rho) = \sup_{\substack{\mathbf{x} \in \mathbf{R}^s \\ r \geq 0}} \{\mu_F(B(\mathbf{x}, r, \rho)) : B(\mathbf{x}, r, \rho) \cap \mathcal{P}_N = \emptyset\},$$

where $B(\mathbf{x}, r, \rho) = \{\mathbf{z} \in \mathbf{R}^s : \rho(\mathbf{z}, \mathbf{x}) \leq r\}$ and $\mu_F(B(\mathbf{x}, r, \rho))$ denotes the probability measure of $B(\mathbf{x}, r, \rho)$, and the supremum is taken over all $\mathbf{x} \in \mathbf{R}^s$ and all radii $r \geq 0$.

We say that the set $\mathcal{P}_N$ of deterministic points is a point set *with distribution* $F(\mathbf{x})$ under the F-volume-dispersion criterion, if

$$\lim_{N\to\infty} \mathcal{V}_F(\mathcal{P}_N, \rho) = 0.$$

Some care must be exercised when speaking about the distribution of a deterministic point set or sequence, since a deterministic point set or sequence does not share certain properties of a random sequence of numbers.

Intuitively, the F-volume-dispersion is the largest probability measure of the ball $B(\mathbf{x}, r, \rho)$, which contains no point of $\mathcal{P}_N$. In this sense, the F-volume-dispersion is a reasonable measure of representation of a point set $\mathcal{P}_N$ for the distribution $F(\mathbf{x})$. When $F(\mathbf{x})$ is the uniform distribution on I^s, then the F-volume-dispersion reduces to the volume-dispersion discussed in Section 2. The F-volume-dispersion is closely related to the F-discrepancy defined below.

DEFINITION 3.2. *The F-discrepancy $D_F(\mathcal{P}_N)$ of $\mathcal{P}_N$ with respect to the distribution $F(\mathbf{x})$ is defined by*

$$D_F(\mathcal{P}_N) = \sup_{J\in E} \left| \frac{A(J;\mathcal{P}_N)}{N} - \mu_F(J) \right|, \tag{11}$$

where E is the family of all subintervals of $\mathbf{R}^s$ of the form $\prod_{i=1}^{s}[u_i, v_i]$ and $\mu_F(J)$ is the probability measure of J.

The next theorem relates the F-volume-dispersion to the F-discrepancy, which is an extension of Theorem 2.2.

THEOREM 3.1. *Letting $\mathcal{P}_N$ be a point set of N points in $\mathbf{R}^s$, then the F-volume-dispersion of $\mathcal{P}_N$ with respect to the maximum metric ρ_∞ is bounded by its F-discrepancy, i.e.,*

$$\mathcal{V}_F(\mathcal{P}_N, \rho_\infty) \le D_F(\mathcal{P}_N). \tag{12}$$

PROOF. Observe that under the maximum metric ρ_∞, for any $\mathbf{x} \in \mathbf{R}^s$ and any $r \ge 0$, the ball $B(\mathbf{x}, r, \rho_\infty)$ is a subinterval in $\mathbf{R}^s$ of the form $\prod_{i=1}^{s}[u_i, v_i]$. Such a form of subintervals is exactly required in the definition of the F-discrepancy. Consider the balls $B(\mathbf{x}, r, \rho_\infty)$ with $B(\mathbf{x}, r, \rho_\infty) \cap \mathcal{P}_N = \emptyset$. For such balls $B(\mathbf{x}, r, \rho_\infty)$,

$$\mu_F(B(\mathbf{x}, r, \rho_\infty)) = \left| \frac{A(B(\mathbf{x}, r, \rho_\infty);\mathcal{P}_N)}{N} - \mu_F(B(\mathbf{x}, r, \rho_\infty)) \right|.$$

By taking supremum over all $\mathbf{x} \in \mathbf{R}^s$ and all radii $r \ge 0$ with $B(\mathbf{x}, r, \rho_\infty) \cap \mathcal{P}_N = \emptyset$, one obtains (12). ∎

So the notion of F-volume-dispersion is weaker than that of F-discrepancy. Any point set (or sequence) with low F-discrepancy has low F-volume-dispersion.

It is then reasonable to ask whether it is possible to generate point sets with a nonuniform distribution. How shall we choose point sets for which the F-volume-dispersion rapidly converges to zero? In general, it is difficult to directly construct a point set with the desired distribution. We will develop methods using low volume-dispersion point sets on I^t to generate low F-volume-dispersion point sets on $\mathbf{R}^s$.

Suppose that the random vector $\mathbf{X} = (X_1, \dots, X_s)$ has a stochastic representation

$$\mathbf{X} = h(\mathbf{U}), \tag{13}$$

where the random vector $\mathbf{U}$ is distributed uniformly on the t-dimensional unit cube I^t, $t \le s$, and h is a continuous function on I^t. Let

$$\mathcal{P}_N^* = \{\mathbf{u}_k \in I^t,\ k = 1, \dots, N\}$$

be a deterministic point set of N points in I^t with uniform distribution. We wish to transform the set $\mathcal{P}_N^*$ into a point set on $\mathbf{R}^s$ with small F-volume dispersion with respect to $F(\mathbf{x})$. A natural way is to define

$$\mathcal{P}_N = \{\mathbf{x}_k = h(\mathbf{u}_k) : \mathbf{u}_k \in \mathcal{P}_N^*,\ k = 1, \dots, N\}. \tag{14}$$

If the set $\mathcal{P}_N^*$ has low volume-dispersion, then we could expect that the corresponding set $\mathcal{P}_N$ of points on $\mathbf{R}^s$ forms a "good" representation for the distribution $F(\mathbf{x})$ in some well-defined sense. We introduce the concept of "quasi-F-volume-dispersion". For each $r \geq 0$ and $\mathbf{v} \in I^t$, let

$$G(\mathbf{v}, r) = \left\{\mathbf{x} \in \mathbf{R}^s : \mathbf{x} = h(\mathbf{u}),\ \mathbf{u} \in B\left(\mathbf{v}, r, \rho^{(t)}\right)\right\},$$

where $\rho^{(t)}$ is a metric on I^t and $B(\mathbf{v}, r, \rho^{(t)}) = \{\mathbf{u} \in I^t : \rho^{(t)}(\mathbf{u}, \mathbf{v}) \leq r\}$ is a ball in I^t of radius r centered at the point $\mathbf{v}$. The set $G(\mathbf{v}, r) \subset \mathbf{R}^s$ can be viewed as a "quasi-ball". The transformation (13) maps the ball $B(\mathbf{v}, r, \rho^{(t)})$ in I^t to the "quasi-ball" $G(\mathbf{v}, r)$ in $\mathbf{R}^s$.

DEFINITION 3.3. *The quasi-F-volume-dispersion $\mathcal{V}_{QF}(\mathcal{P}_N, \rho^{(t)})$ of the point set $\mathcal{P}_N$ with respect to $F(\mathbf{x})$ is defined by*

$$\mathcal{V}_{QF}\left(\mathcal{P}_N, \rho^{(t)}\right) = \sup_{\substack{\mathbf{v} \in I^t \\ r \geq 0}} \{\mu_F(G(\mathbf{v}, r)) : G(\mathbf{v}, r) \cap \mathcal{P}_N = \emptyset\},$$

where $\mu_F(G(\mathbf{v}, r))$ denotes the probability measure of the "quasi-ball" $G(\mathbf{v}, r)$.

We will also say that $\mathcal{P}_N$ is a quasi-random point set *with distribution* $F(\mathbf{x})$ under the quasi-F-volume-dispersion criterion, if $\lim_{N\to\infty} \mathcal{V}_{QF}(\mathcal{P}_N, \rho) = 0$.

Intuitively, quasi-F-volume-dispersion is also a reasonable measure of representation of $\mathcal{P}_N$ for $F(\mathbf{x})$, because it is the largest probability measure of the "quasi-ball" $G(\mathbf{v}, r)$, which contains no point of $\mathcal{P}_N$.

THEOREM 3.2. *Let the point set $\mathcal{P}_N$ be generated by (14); then its quasi-F-volume-dispersion is equal to the volume-dispersion of the original point set $\mathcal{P}_N^*$, i.e.,*

$$\mathcal{V}_{QF}\left(\mathcal{P}_N, \rho^{(t)}\right) = \mathcal{V}_d\left(\mathcal{P}_N^*, \rho^{(t)}\right).$$

This can be proved from the following facts: for each $r \geq 0$ and any point $\mathbf{v} \in I^t$, let $\mathbf{x} = h(\mathbf{u})$; then

(1) $\mathbf{x} \in G(\mathbf{v}, r)$ if and only if $\mathbf{u} \in B(\mathbf{v}, r, \rho^{(t)})$;
(2) $G(\mathbf{v}, r) \cap \mathcal{P}_N = \emptyset$ if and only if $B(\mathbf{v}, r, \rho^{(t)}) \cap \mathcal{P}_N^* = \emptyset$;
(3) $\mu_F(G(\mathbf{v}, r)) = \mathrm{Vol}\,[B(\mathbf{v}, r, \rho^{(t)})]$.

Thus, starting from a set $\mathcal{P}_N^*$ of N points in I^t with small volume-dispersion and using (14), one can obtain a set $\mathcal{P}_N$ of N points in $\mathbf{R}^s$ with small quasi-F-volume-dispersion. In this sense, the set $\mathcal{P}_N$ forms a good representation for $F(\mathbf{x})$.

Theorem 3.2 provides a theoretical basis for using some random sampling techniques in Monte Carlo methods to generate low quasi-F-volume-dispersion point set with the desired probability distribution. There are a host of methods for generating random variates in Monte Carlo context. See, for example, [11,12]. The inverse transform method and the conditional distribution method are the most general methods for generating nonuniform random variates. Based on Theorem 3.2, the deterministic versions of these methods can be used to generate low quasi-F-volume-dispersion point sets with the desired distribution, since both of these methods are some kind of transformation. The point sets obtained by using such techniques are reasonably good under quasi-F-volume-dispersion criterion, as indicated by the following results.

COROLLARY 3.3. *Let $F(\mathbf{x})$ be an s-dimensional continuous distribution function defined on $\mathbf{R}^s$ with independent marginal $F_i(x_i)$*

$$F(\mathbf{x}) = F(x_1, \dots, x_s) = \prod_{i=1}^{s} F_i(x_i).$$

If $\mathcal{P}_N^* = \{\mathbf{u}_k = (u_{k1}, \dots, u_{ks}),\ k = 1, \dots, N\}$ is a set of N points on I^s with volume-dispersion $\mathcal{V}_d(\mathcal{P}_N^*, \rho)$, then the set on $\mathbf{R}^s$

$$\left\{\left(F_1^{-1}(u_{k1}), \dots, F_s^{-1}(u_{ks})\right),\ k = 1, \dots, N\right\}$$

has quasi-F-volume-dispersion equal to $\mathcal{V}_d(\mathcal{P}_N^*, \rho)$.

COROLLARY 3.4. *Let the random vector $\mathbf{X} = (X_1, \dots, X_s)$ have distribution function $F(\mathbf{x})$. Let $\mathcal{P}_N^*$ be the same as in Corollary 3.3. Suppose $F_1(x_1)$ is the c.d.f. of X_1, and $F_i(x_i \mid x_1, \dots, x_{i-1})$ is the conditional c.d.f. of X_i given $X_j = x_j$, $j = 1, \dots, i-1$; $i = 2, \dots, s$. Let $\mathbf{x}_k = (x_{k1}, \dots, x_{ks})$ be the solutions of the following system of equations for $k = 1, 2, \dots, N$:*

$$\begin{aligned} F_1(x_{k1}) &= u_{k1}, \\ F_i(x_{ki} \mid x_{k1}, \dots, x_{k,i-1}) &= u_{ki}, \qquad i = 2, \dots, s. \end{aligned}$$

Then the set $\{\mathbf{x}_k,\ k = 1, \dots, N\}$ has quasi-F-volume-dispersion equal to $\mathcal{V}_d(\mathcal{P}_N^*, \rho)$.

These results indicate that the quality of the point sets generated by inverse transformation method or by conditional transformation method depends only on the underlying point set $\mathcal{P}_N^*$. If the point set $\mathcal{P}_N^*$ has volume-dispersion of the optimal order $O(N^{-1})$, then the obtained set $\{\mathbf{x}_k,\ k = 1, \dots, N\}$ has quasi-F-volume-dispersion also in the order $O(N^{-1})$.

In some cases, the inverse transformation method or conditional transformation method is difficult, or even impossible, to be used due to the complexity of the probability distribution functions. In such settings, the deterministic version of *acceptance-rejection method* may be used. This method in QMC setting for numerical integration is studied in [13] and indicates its suitability for generating deterministic points for numerical integration. Moreover, it is shown there that the F- discrepancy of the point set generated by the deterministic version of acceptance-rejection method (using low discrepancy sequences) is smaller than the F-discrepancy of the point set generated by the random version of acceptance-rejection method (using random numbers). According to Theorem 3.1, a point set with small F-discrepancy has small F-volume-dispersion. Thus, the deterministic version of acceptance-rejection method can also be used to generate a low F-volume-dispersion point set with the desired distribution. Various variations of acceptance-rejection method in Monte Carlo context [12] are also possible in a deterministic setting.

Now consider briefly the application of quasi-F-volume-dispersion in the problems of global optimization. Let $F(\mathbf{x})$ be a given probability distribution function. Often this distribution reflects some prior information of the location of optimal point or some constraints on the optimization problems. Let $\mathcal{P}_N = \{\mathbf{x}_k,\ k = 1, \dots, N\}$ be a point set aiming for low F-volume-dispersion or low quasi-F-volume-dispersion with respect to $F(\mathbf{x})$. How can the optimization approximation error $f^* - f_N^*$ be bounded? Here, as before, $f^* = \sup_{\mathbf{x} \in \mathbf{R}^s} f(\mathbf{x})$ and $f_N^* = \max_{1 \le i \le N} f(\mathbf{x}_i)$.

Suppose that the corresponding random vector $\mathbf{X}$ has a relationship (13) with the uniformly distributed random vector $\mathbf{U}$. Let $g(\mathbf{u}) = f(h(\mathbf{u}))$. Then the global optimization problem $f^* = \sup_{\mathbf{x} \in \mathbf{R}^s} f(\mathbf{x})$ is equivalent to $g^* = \sup_{\mathbf{u} \in I^t} g(\mathbf{u})$. Let the point set $\mathcal{P}_N$ be generated by (14), starting from a low volume-dispersion point set P_N^* on I^t. Denote $g_N^* = \max_{1 \le i \le N} g(\mathbf{u}_i)$. Note that $f(\mathbf{x}_i) = g(\mathbf{u}_i)$ for all $i = 1, \dots, N$. Therefore,

$$f^* - f_N^* = g^* - g_N^*.$$

As shown in Section 2, the right-hand side can be bounded in terms of the volume-dispersion of the point set $\mathcal{P}_N^*$ (see Theorem 2.4), which is the same as the quasi-F-volume-dispersion of the point set $\mathcal{P}_N$ by Theorem 3.2. This means that the optimization approximation error $f^* - f_N^*$ can be bounded in terms of the quasi-F-volume-dispersion of the set $\mathcal{P}_N$ of searching points. In this sense, the quasi-F-volume-dispersion can be viewed as a measure for the effectiveness of $\mathcal{P}_N$ in QMC search. This shows the usefulness of the measure of quasi-F-volume-dispersion in the problems of global optimization. An advantage of QMC search based on the quasi-F-volume-dispersion criterion other than on the traditional dispersion is that one can cluster points near to where the maximum is assumed to be. This can enhance the efficiency of the QMC search algorithms.

In this section, we have assumed that the probability distribution $F(\mathbf{x})$ is known. In practice, we may not know it. We can estimate $F(\mathbf{x})$ from the information gained from a previous search using some statistical methods. Based on such an idea, *adaptive* QMC search procedure can be developed. This approach and more applications of F-volume-dispersion will be addressed elsewhere.

4. CONCLUDING REMARKS

Measures of irregularity of distribution play a central pole in QMC methods for choosing good point sets. The newly defined volume-dispersion and F-volume-dispersion (including quasi-F-volume-dispersion) are closely related to the well-known discrepancies and traditional dispersion. The volume-dispersion has good properties and has some advantages over discrepancies and traditional dispersion. The generalizations to F-volume-dispersion and quasi-F-volume-dispersion show the increased power of the new measure of irregularity of distribution compared with the traditional dispersion. Many works remain to be done with volume-dispersion. We have not succeeded in construction of an explicit formula for the evaluation of volume-dispersion for arbitrary point set in dimension $s \geq 2$. The applications of low volume-dispersion and low (quasi-)F-volume-dispersion point sets in optimization problems with constraints, as well as in the problems of experimental design, are also worth studying and are the topics of ongoing research. Note that in the problem of experimental design, the traditional dispersion criterion leads to the well-known *minimax distance design* [14]. An N points design is called a minimax distance design, if it minimizes the traditional dispersion among all possible N points designs. It is shown that minimax distance designs have quite asymptotically general optimum in suitable problems [14]. Similarly, we can define a design of N points to be "minimum (F-) volume-dispersion design", if it minimizes the (F-) volume-dispersion among all possible N points designs. We believe that the volume-dispersion and the F-volume-dispersion criteria will also lead to interesting experimental designs.

REFERENCES

1. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, PA, (1992).
2. K.T. Fang and Y. Wang, *Number-Theoretic Methods in Statistics*, Chapman & Hall, London, (1994).
3. J.H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* **2**, 84–90, (1960).
4. I.M. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, *Zh. Vychisli. Mat. i Mat. Fiz.* **7**, 784–802, (1967).
5. H. Faure, Discrépance de suites associées à un système de numération (en dimension s), *Acta Arith.* **41**, 337–351, (1982).
6. R.Y. Rubinstein, *Monte Carlo Optimization, Simulation and Sensitivity of Queuing Networks*, Wiley, New York, (1986).
7. H. Niederreiter, A quasi-Monte Carlo method for approximate evaluation of the extreme value of a function, In *Studies in Pure Mathematics*, pp. 523–529, Birkhäuser, Basel, (1983).
8. I.H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, *J. Complexity* **14**, 1–33, (1998).

9. S.H. Paskov and J.F. Traub, Faster valuation of financial derivatives, *J. Portfolio Management* **22**, 113–120, (1995).
10. I.M. Sobol', On the crude multidimensional search, *J. Comput. Appl. Math.* **56**, 283–293, (1994).
11. L. Devroye, *Non-Uniform Random Variates Generation*, Springer-Verlag, New York, (1986).
12. J.E. Gentle, *Random Number Generation and Monte Carlo Methods*, Springer, New York, (1998).
13. X. Wang, Improving the rejection sampling methods in quasi-Monte Carlo methods, *J. Comput. Appl. Math.* **114**, 231–246, (2000).
14. M.E. Johnson, L.M. Moore and D. Ylvisaker, Minimax and maxmin distance designs, *J. Statis. Plann. Inference* **26**, 131–148, (1990).